AN APPLICATION OF BAYESIAN STATISTICAL METHODS IN THE DETERMINATION OF SAMPLE SIZE FOR OPERATIONAL TESTING IN THE U.S. ARMY

A THESIS

Presented to

The Faculty of the Division of Graduate Studies

by

Robert M. Baker

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Operations Research

Georgia Institute of Technology

June, 1976

AN APPLICATION OF BAYESIAN STATISTICAL METHODS

IN THE DETERMINATION OF SAMPLE SIZE

FOR OPERATIONAL TESTING IN THE U.S. ARMY

Appr

Douglas C. Montgomery, Chairman

Lynwood A. Johnson

Richard D. Wright

Date approved by Chairman: 5/31/76

ACKNOWLEDGMENTS

I would like to express my appreciation to my thesis advisor, Dr. Douglas C. Montgomery, for his encouragement and understanding in assisting me in the preparation of this thesis. My gratitude is also extended to Dr. Lynwood A. Johnson and Dr. Richard D. Wright for their many significant comments and useful suggestions.

A special note of thanks goes to my wife, Mary Frances, for her enduring patience and endless encouragement over the past two years.

TABLE OF CONTENTS

TABLE OF CONTENTS (Concluded)

LIST OF TABLES

LIST OF ILLUSTRATIONS

SUMMARY

This research is devoted to investigating how Bayesian statistical procedures might be used to improve the design of operational tests being conducted by the U.S. Army Operational Test and Evaluation Agency. The specific aspect of the design process which is of concern is the calculation of required sample sizes. Basically, three changes are suggested in the methods currently being employed.

First, it is shown that the problem can be reformulated in a manner which is believed to be more closely aligned with the objectives of operational testing, in so doing, it is possible to capitalize on the comparative nature of the testing.

The problem is then analyzed using Bayes' theorem and Bayesian inference techniques. It is felt that the application of Bayes' theorem can provide for a more efficient use of information available to test design personnel and that this may result in a reduction in required sample sizes when compared to methods presently being utilized. Formulas are then derived for calculating the sample size required to reduce the expected value of selected measures of tightness of the posterior distribution.

Finally, a method is proposed for utilizing these procedures in the presence of economic considerations such as budget constraints and sampling costs. This method attempts to find the economically optimal sample size by systematically comparing the cost of experimentation with the value of the information expected to be obtained from that experimentation.

CHAPTER I

INTRODUCTION

Background

The impetus for this study was provided by the interest of the U.S. Army Operational Test and Evaluation Agency (OTEA) to investigate the possible application of Bayesian statistical analysis and decision theory to sample size determination for operational testing. In order to understand some of the procedures discussed later in this study, a basic knowledge of the nature of operational testing as performed by OTEA is necessary. The purpose of operational testing is to provide a source of data from which estimates may be developed as to the military utility, operational effectiveness and operational suitability of new weapon systems. This data is obtained through a sequence of three operational tests; each test in the sequence is completed and the results analyzed prior to beginning the next test. For ease of reference, these tests will be referred to as Operational Test I (OT I), Operational Test II (OT II) and Operational Test III (OT III). Once the data has been collected and the estimates developed an assessment is made of the new system's desirability as compared to systems which are already available [2].

The overall assessment procedure parallels closely that proposed by Miller in his book, Professional Decision Making [10]. Initially, certain issues concerning the system's capability are identified for further examination before any assessment of the overall desirability is made. These issues are general in nature and cannot be resolved directly. For example, a system's fire power may be an issue of interest. This cannot be represented by any single, physically measurable quantity. Therefore, it is necessary to refine the issues into a number of parameters which adequately represent the issue and which can be measured quantitatively. These quantities are known in OTEA as measures of effectiveness (MOE). For the example given above, the MOE might include such things as percent of targets hit, mean miss distance, percent of fire requests which are met, proportions of rounds requested fired and so on [11]. Once these MOE are identified, an operational test is designed to provide for a side-by-side comparison of the competing systems with respect to each MOE.

Given a fixed test design, the sample size problem becomes one of determining the minimum number of replicates required for each set of experimental conditions in order to produce sufficient sample information upon which to base statistically valid inferences. This problem can become quite complex since a single operational test may involve as many as a hundred MOE.

An approach which has been recommended to reduce the computational burden is to rank-order the MOE based on their relative importance and then to calculate the sample size requirements using only two or three of the more important MOE. These calculations are presently based on classical statistical procedures [11].

Objectives of Research

In reviewing these procedures, two areas of possible improvement were identified. The first is concerned with making efficient use of all available data. As noted earlier, the operational testing program is sequential in nature and, many times, the same measure of effectiveness may be examined in more than one test. When this occurs, the data from the previous test is sometimes used in the design of the subsequent test in that it serves as a basis for the formulation of hypotheses and as a source of variance estimates for sample size calculations. This data is not, however, being combined with the data obtained during later tests in the final statistical analysis. By not doing this, it is felt that valuable information is being wasted. In fact, it is believed that, if this information were used to its fullest extent, a reduction in the required sample size would be possible. One method of combining prior information with sample results is provided by Bayes' theorem. The next chapter is devoted to investigating how Bayes' theorem might be applied in the operational testing environment and what effect this would have on the calculation of required sample sizes.

The second area identified for possible improvement is concerned with the economics involved in experimentation. Presently the costs associated with proposed experiments are not directly considered in sample size calculations. Additionally, there is no evidence of a quantitative assessment of the expected value of the sample information to be obtained from a particular experiment. Considering this, it is doubtful that the money available for testing is being allocated to the various experiments in an optimal fashion. It is felt that this problem might best be analyzed using the concepts of Bayesian statistical decision theory. In this manner, the economics of testing could be considered explicitly and the economically optimal sample size could be determined for each experiment. In Chapter III, the application of these ideas to operational testing will be examined.

Fundamentals of Bayesian Analysis

Because OTEA is currently using classical statistical methods, the discussion presented here will be comparative in nature. That is, the Bayesian ideas will be contrasted with classical statistical ideas, and similarities and differences highlighted. From the outset, there are some fundamental conceptual differences requiring discussion. Consider the situation in which a particular data-generating process may be modeled by the normal process with unknown mean and variance. Then the probability density function associated with such a process is the normal density with mean, μ, and variance, σ^2. The classical statistician would view these parameters as unknown constants. He might decide to estimate them by taking a sample from the data-generating process (or an appropriate model thereof) and use the sample statics $\bar{x}$ and s^2 as estimates of μ and σ^2, respectively. If he is interested in constructing a confidence interval on x, he could substitute these estimates into the normal density function making it possible to compute the probability that a particular observation would lie within a specified interval, i.e., $P(x_1 \leq x \leq x_2) = p$. This probability would then be interpreted in the relative frequency sense. That is, if a large number of observations were taken it would be expected that x would lie on the interval (x_1, x_2), "p" percent of the time [8]. On the other hand, the Bayesian analyst would view the unknown parameters, $\tilde{\mu}$ and $\tilde{\sigma}^2$, as random variables (throughout this paper, a "tilde" will be used to denote a random variable). As such, he would not attempt initially to obtain a point estimate of these parameters. Instead, he would ascribe to them a probability distribution. Prior to sampling from the process, such a probability distribution must be constructed based on the analyst's prior beliefs concerning the joint occurrence of $\tilde{\mu}$ and $\tilde{\sigma}^2$. A probability distribution constructed in this manner reflects the analyst's subjective probabilities on $\tilde{\mu}$ and $\tilde{\sigma}^2$. It will be shown later how these probabilities may be combined with sample information to produce new distributions on the unknown parameters. The conceptual differences discussed here play important roles in interpreting the results of an analysis [23].

As mentioned above, Bayesian analysis can be used to combine sample information with prior beliefs in an effort to develop a probability distribution for a random variable. This combination is achieved by using Bayes' theorem. For a continuous random variable, say $\tilde{\theta}$, Bayes' theorem may be written as

$$f''(\theta|y) = \frac{f'(\theta)f(y|\theta)}{\int_{-\infty}^{\infty} f'(\theta)f(y|\theta)d\theta} \tag{1-1}$$

In this notation*, a single prime superscript (') denotes a prior distribution or parameter, a double prime (") denotes a posterior distribution

*The notation used in this study is similar to that used by Raiffa and Schlaifer [18]. At times it can become quite intricate; therefore, a detailed explanation of this notation is presented in Appendix III.

or parameter and no superscript designates a sampling distribution or parameter. Therefore, in equation (1-1), $f'(\theta)$ is the prior distribution of $\tilde{\theta}$ representing the decision maker's beliefs regarding $\tilde{\theta}$ prior to sampling, $f(y|\theta)$ represents the likelihood function chosen to describe the sampling process and $f''(\theta|y)$ is the posterior distribution of $\tilde{\theta}$ which reflects the decision maker's beliefs regarding $\tilde{\theta}$ after the sample has been taken [23]. An analogous form of the theorem may be written for discrete random variables by substituting probability mass functions for the probability density functions and a summation sign for the integral sign. A derivation of Bayes' theorem from conditional probability formulas is given by Winkler [23]. In the application of Bayes' theorem the major difficulties lie in the assessment of the prior distribution and the likelihood function and, in the continuous case, in the evaluation of the integral appearing in the denominator of the formula. Suggested methods for handling these difficulties are discussed in the next chapters.

CHAPTER II

AN INFERENTIAL APPROACH

Bayesian Inference

Statistical inference is the process of forming reasonable conclusions about some aspect of a random phenomenon. For example, in considering the mean, μ, of a normal distribution, the classical statistician may attempt to estimate the true value of μ based on sample information. Alternatively, he may construct a confidence interval on μ of the form described in Chapter I. The Bayesian statistician also makes inferential statements regarding $\tilde{\mu}$. These statements, however, are developed in a different manner and have different interpretations than their classical counterparts.

As pointed out in Chapter I, the Bayesian considers $\tilde{\mu}$ to be a random variable and assigns to it some probability distribution. Inferential statements concerning $\tilde{\mu}$ are then based on this distribution. For instance, while the classical statistician may estimate the true value of $\tilde{\mu}$, the Bayesian may be interested in an estimate of the most likely value of $\tilde{\mu}$ and may use as a point estimate the mode of the distribution [23]. In making interval estimates of $\tilde{\mu}$, the Bayesian attempts to define an interval [a,b] such that the probability that $\tilde{\mu}$ will take on values between a and b is some number "p."

Thus, Bayesian inference revolves around the distribution of the unknown quantity of interest. If the analyst's present state of knowledge about this unknown quantity is sufficient to develop a well-defined prior distribution, then this distribution may be used for the analysis. On the other hand, if the analyst's prior knowledge is vague, he may consider gathering sample information in order to produce a posterior distribution upon which to base his inferences. In this case, the problem is to determine how much sample information is necessary to produce a suitable posterior distribution. Two possible approaches to solving this problem are presented later in this chapter. First, however, the two required inputs to Bayes' theorem, i.e., the likelihood function and the prior distribution, will be discussed in the context of operational testing.

The Likelihood Function

In operational testing, the value of the MOE under consideration may be thought of as the uncertain state of nature and may be represented by $\tilde{\theta}$. If y is a sufficient statistic for a sample from the data-generating process, then prior to sampling, $\tilde{y}$ is also a random variable. The probability distribution of y is assumed to depend on θ, and the conditioned probability distribution of y given θ will be denoted by $f(y|\theta)$ and called the "likelihood" function. In order to proceed with the analysis, it is necessary to mathematically describe this function. In doing this, subjective probability assessments could be made for each data-generating process encountered and unique likelihood functions constructed. This could prove extremely cumbersome considering the number of MOE involved in a single operational test. A better approach, and one used more frequently in practice, is to attempt to "fit" one of the more common statistical models to the process. As pointed out by Winkler [23], this does not eliminate the subjectivity involved in assessing the likelihood function, although it may make it somewhat less controversial.

In choosing a likelihood function to be used in this study, it was desired to select a function which would realistically represent a broad class of MOE. The univariate normal with unknown mean and variance is such a function. Its applicability to a wide range of MOE is supported by the fact that it is currently being used by OTEA as the basic model for sample size determination for both measurement and attribute data [11]. It should be stressed, however, that this likelihood function should not be used indiscriminately but only when the decision maker's prior beliefs concerning the data-generating function suggest that it would be an appropriate model.

The Prior Distribution

Before any operational testing is conducted, the prior distribution would have to be assessed based on the decision maker's prior notions concerning $\tilde{\theta}$ or, in the case of a totally informationless situation, would have to be represented by a diffuse distribution, which will be discussed later. After at least one operational test has been conducted, these sample results might be used in constructing the priors for similar MOE in later tests [11].

In selecting a prior distribution, several desirable characteristics should be considered. First, and most important, the distribution should adequately reflect the decision maker's prior beliefs. Second, it should be of such a form as to be mathematically tractable when combined with the likelihood function in Bayes' theorem. And, finally, it would be desirable to have the resulting posterior distribution be of the same form as the prior so as to reduce the computational burden in sequential analysis.

By choosing a prior distribution from the natural conjugate family of distributions it is possible to guarantee that the second and third conditions above are met (for a definition and complete discussion of natural conjugate distributions see Raiffa and Schlaifer [18], Chapter 3). While the first condition is not guaranteed to be satisfied, it is provided for by the fact that natural conjugate families are generally "rich" and through proper parameterization can be made to represent a wide variety of distributions.

The choice of a natural conjugate family is determined by the form of the likelihood function. Since the likelihood function used here is normal with unknown mean and variance, the appropriate conjugate family is the normal-gamma distribution [18]. Which member of this family, if any, will be appropriate depends on the amount and substance of the decision maker's prior beliefs concerning the specific MOE under consideration. Certain peculiarities of operational testing will allow for some general statements to be made concerning the decision maker's prior state of knowledge. First is the requirement, imposed by the Department of the Army, that operational testing be independent of all other testing. This severely limits the use of any prior knowledge on $\tilde{\theta}$. For all practical purposes, prior to OT I there exists a self-imposed, totally informationless situation. In such cases, the prior information, or lack thereof, should be represented by a diffuse prior distribution.

In general, a diffuse distribution need not be limited in its use to a totally informationless situation, but may be used whenever the decision maker's prior information is diffuse relative to that which can be obtained through sampling. As a rule, diffuse distributions are characterized by large variances as compared to that of the data-generating function, they are relatively flat in the region where the likelihood function takes on significant values, and they are given little weight in comparison to the sampling distribution when computing the posterior distribution [23]. The real objective in selecting a diffuse distribution is to choose one which will have no effect on the posterior distribution [23]. With this objective in mind, the actual form of a diffuse prior is of little importance, in fact, it need not even be a proper probability distribution [23]. Thus, it is only rational to choose a diffuse distribution from the family of natural conjugate distributions, in this case, the normal-gamma family. This family of distributions is of the following form [18],

$$f_{N\gamma}(\tilde{\mu},\tilde{h}|m',v',n',\nu') \propto e^{-\frac{1}{2}\tilde{h}n'(\tilde{\mu}-m')^2}\,\tilde{h}^{\frac{1}{2}}\,e^{-\frac{1}{2}\tilde{h}\nu'v'}\,\tilde{h}^{\frac{1}{2}\nu'-1}$$

where $\tilde{\mu}$ and $\tilde{h}$ are random variables such that

$$-\infty < \tilde{\mu} < \infty$$

$$\tilde{h} > 0$$

and m', v', n', ν' are parameters such that

$$v', n', \nu' > 0$$

$$-\infty < m' < \infty$$

The above parameters may be interpreted as previous sample results associated with some actual or hypothetical experiment [23]. Using this interpretation, m' would be equivalent to the sample mean, v' would be the sample variance, n' the size of the previous sample and ν' equal to n' minus one (usually called the number of degrees of freedom for v').

By the properties of natural conjugate distributions, it is known that the posterior distribution will be of the same form. Raiffa and Schlaifer [18] have shown that the posterior parameter (m'',v'',n'',ν'') is given by

$$m'' = \frac{n'm' + nm}{n' + n} \tag{2-1}$$

$$n'' = n' + n$$

$$v'' = \frac{\nu'v' + n'm'^2 + \nu v + nm^2 - n''m''^2}{\nu' + \delta(n') + \nu + \delta(n) - \delta(n'')}$$

$$\nu'' = \nu' + \delta(n') + \nu + \delta(n) - \delta(n'')$$

where $\delta(\gamma)$ is an indicator variable defined by

$$\delta(\gamma) = \begin{cases} 1 \text{ if } \gamma > 0 \\ 0 \text{ if } \gamma = 0 \end{cases}$$

and (m,v,n,ν) is the statistic resulting from a sample of size n and is given by

$$m = \frac{\Sigma X_i}{n}$$

$$v = \frac{\Sigma(X_i - m)^2}{n-1}$$

$$\nu = n-1$$

It is shown in reference 18, p. 300 that, if $n'=\nu'=0$, the posterior parameter (m'',v'',n'',ν'') equals the sampling statistic (m,v,n,ν). Therefore, any normal-gamma distribution can be used as a diffuse prior so long as $n'=\nu'=0$.

Now, consider the case where at least one operational test has already been conducted. This will provide usable data on a number of MOE. Commonly, several of these MOE will be subject to further examination in subsequent operational tests. Under these circumstances the posterior distribution of the earlier test may be used as the prior distribution for the later test. This is especially useful if the distributions involved are natural conjugates of the likelihood function.

Even if the MOE to be evaluated in later tests have not been examined previously, they may possess strong similarities to MOE which have. In such cases, it might be possible to construct prior distributions based on subjective probability notions derived from data on the earlier MOE.

In summary, if there is no internally-generated data available from which to develop prior distributions, a diffuse normal-gamma prior will be used. If the available data is of the form of a previous sample from the same data-generating process, then the posterior distribution of the earlier test will be used as the prior for the later test. Finally, if the available data is of the form of a sample from a similar data-generating process, then this posterior distribution will be used in making a subjective assessment of the prior distribution to be used in the later test.

The General Problem

As noted in Chapter I, operational tests are designed to provide a side-by-side comparison of competing weapons systems. The overall objective of this testing is not so much to estimate the performance characteristics of either of the systems, but rather to make inferences regarding the difference in these performance characteristics. In this context, the observations may be considered to be paired observations and the difference in the observations may be viewed as a random variable with its own probability distribution. Consider, as a hypothetical example, the problem of determining whether a new weapons system has a greater range than that of the existing system which it has been designed to replace. Let the range of the existing system be denoted by $\tilde{X}_1$ and that of the new system by $\tilde{X}_2$. Assume that the prior information on $\tilde{X}_1$ and $\tilde{X}_2$ is such that both may be modeled by the normal process with unknown mean and variance. Then, the difference in the range of the two systems is also a random variable, $\tilde{D}$, given by

$$\tilde{D} = \tilde{X}_1 - \tilde{X}_2 \tag{2-2}$$

From equation (2-2), $\tilde{D}$ is merely the linear combination of two independent, normally distributed random variables, which implies that $\tilde{D}$ is also normally distributed [8] with unknown mean and variance, say $\tilde{\mu}$ and $\tilde{\sigma}^2$, respectively. Using this distribution of $\tilde{D}$, the family of natural conjugate prior distributions is normal-gamma. The mathematics involved in working with this particular family of distributions can be quite complex, fortunately, this may not be necessary. What is of particular importance in this testing is the mean difference, $\tilde{\mu}$, between the two systems. Since $\tilde{\mu}$ is itself a random variable, it follows a particular probability distribution. It has been shown that this distribution is Student's t distribution and can be represented by the density [18]

$$f(\tilde{\mu}|m,v,n,\nu) = f_S(\tilde{\mu}|m,n/v,\nu) \tag{2-3}$$

Formulated in this manner, the problem becomes one of determining the minimum sample size which can be expected to produce a posterior distribution suitable for making meaningful probability statements regarding $\tilde{\mu}$. One approach to solving this problem is to identify some measure on the posterior distribution which is a function of the sample size, derive the expected value of this measure, equate this expected value to some desirable value and solve for the sample size. Solution procedures utilizing two such measures have been developed and are presented in the next two sections.

The Solution Using the Standard Deviation

Given that $\tilde{\mu}$ has the density described by equation (2-3), then [18]

$$E(\tilde{\mu}|m,v,n,\nu) \equiv \bar{\mu} = m\ , \qquad \nu > 1$$

$$V(\tilde{\mu}|m,v,n,\nu) \equiv \check{\mu} = \frac{v}{n}\,\frac{\nu}{\nu-2}\ , \qquad \nu > 2$$

Under the prior distribution of $\tilde{\mu}$, the value of $\check{\mu}'$ is given by

$$\check{\mu}' = \frac{v'}{n'}\,\frac{\nu'}{\nu'-2}\ , \qquad \nu' > 2 \tag{2-4}$$

This value may then be used to calculate the prior standard deviation of $\tilde{\mu}$, i.e., $\sqrt{\breve{\mu}'}$. If the prior standard deviation is large relative to $\bar{\mu}'$, then the prior distribution may not be "tight" enough to reach any meaningful conclusions about $\tilde{\mu}$. In this situation, it may be necessary to obtain additional information about $\tilde{\mu}$ through sampling. The objective of this sampling would be to produce a value for the posterior standard deviation which would be small enough to allow inferences to be made about $\tilde{\mu}$.

Suppose that it is felt a posterior standard deviation equal to some specific fraction of the prior standard deviation would be satisfactory. Mathematically, this relationship is

$$\sqrt{\breve{\mu}''} = s\sqrt{\breve{\mu}'} \; , \qquad 0 < s \leqq 1 \; . \tag{2-5}$$

Prior to sampling, the posterior standard deviation is a random variable and, therefore, it is necessary to think in terms of its expected value. This has been shown to be [18]

$$E[\sqrt{\breve{\mu}''} \,|m',v',n',\nu';n,\nu) = \sqrt{(n'/n'')\bar{\mu}'}\, e^{-3/8((1/1/2\ \nu'-1)-(1/1/2\ \nu''-1))} \; . \tag{2-5a}$$

In deriving the above equation, it was necessary to use Stirling's <u>second</u> approximation for the following two values

$$(1/2\ \nu'-1)! \approx (2\pi)^{1/2}(1/2\ \nu'-1)^{1/2(\nu'-1)}e^{-(1/2\ \nu'-1)+(1/(12(1/2\ \nu'-1)))} \tag{2-5b}$$

$$(1/2\ \nu''-1)! \approx (2\pi)^{1/2}(1/2\ \nu''-1)^{1/2(\nu''-1)}e^{-(1/2\ \nu''-1)+(1/(12(1/2\ \nu''-1)))} \; . \tag{2-5c}$$

It has been suggested [18,p. 308] that, if Stirling's <u>first</u> approximation were used, the second term in the exponent of "e," in both equations (2-5b) and (2-5c), would be omitted and the expected value would then become

$$E[\sqrt{\tilde{\mu}''}\,|\,m',v',n',\nu';n,\nu] = \sqrt{(n'/n'')\mu'} \ . \qquad (2\text{-}6)$$

Thus, in using Stirling's first approximation, some information is sacrificed in order to obtain a less complex mathematical expression for the expected value of the posterior standard deviation. The question then becomes one of how much information is lost and is this loss justified.

A problem similar to this is encountered in classical statistics when attempting to arrive at an unbiased estimator of the standard deviation. One solution to this problem is to multiply the sample standard deviation by an appropriate correction factor. Gurland and Tripathi have shown that this correction factor approaches one as the size of the sample increases.* In fact, for a sample of size 20, the correction factor is 1.0132, implying that the sample standard deviation varies from the unbiased estimator of the population standard deviation by only slightly more than one percent.

It is felt that the problem of approximating equation (2-5a) by equation (2-6) may be viewed in a similar fashion where the exponential term in equation (2-5a) is analogous to the correction factor discussed above. Therefore, the percent error induced by the approximation can be

*Gurland, J., and Tripathi, R. C., "A Simple Approximation for Unbiased Estimation of the Standard Deviation," <u>The American Statistician</u>, Vol. 25, 1971, pp. 30-32.

expressed as

$$\%\ \text{error} = 1 - e^{-3/8((1/1/2\ \nu'-1)\ -\ (1/1/2\ \nu''-1))}. \qquad (2\text{-}6a)$$

Note that, using the relationships [18]

$$\nu'' = \nu' + \nu + 1$$

and

$$\nu = n - 1\ ,$$

equation (2-6a) may be rewritten as

$$\begin{aligned}\%\ \text{error} &= 1 - e^{-3/8((1/1/2\ \nu'-1)\ -\ (1/1/2(\nu'+n)-1))} \\ &= 1 - e^{-6/8((1/\nu'-2)\ -\ (1/\nu'+n-2))}.\end{aligned}$$

This equation was used to calculate the percent error of approximation for selected values of n and ν', and the results are presented in Table 1. Notice that, for a given value of ν', the accuracy of the approximation decreases to a limit as the sample size approaches infinity. This phenomenon makes it possible to establish an upper bound on the approximation error for any given value of ν'. For the values of ν' considered in Table 1, this upper bound is shown in the last row of the table. From this information it can be seen that the approximation is reasonably accurate for values of ν' greater than or equal to 35 regardless of the size of the sample. For values of ν' less than 35, the decision about whether or not to use the approximation would have to be made based on a comparison of the percent error induced in the calculations versus the desired accuracy of the results.

Table 1. Percent Error in Approximating the Expected Posterior Standard Deviation

$n \backslash \nu'$	5	10	15	20	25	30	35	40	45	50
5	14	4	2	1	1	0	0	0	0	0
10	17	5	2	1	1	1	1	0	0	0
15	19	6	3	2	1	1	1	1	0	0
20	19	6	3	2	2	1	1	1	1	0
25	20	7	4	2	2	1	1	1	1	1
30	20	7	4	3	2	1	1	1	1	1
35	20	7	4	3	2	1	1	1	1	1
40	21	8	4	3	2	2	1	1	1	1
45	21	8	4	3	2	2	1	1	1	1
∞	22	9	6	4	3	3	2	2	2	2

The procedures developed in the remainder of this study utilize the approximate expression for the expected value of the posterior standard deviation. It is believed that this approximation will be acceptable in the design of a large percentage of operational tests. When it is not acceptable, the methodology presented in this study might still be applicable, although the specific results would not be. For example, it may be possible to use the exact expression given by equation (2-5a). Although this equation cannot be solved explicitly for n, it can be solved iteratively. In the iterative solution, it is suggested that the analyst use equation (2-6) to calculate a first approximation for n.

Returning to the development of the methodology, if equation (2-6) is used in place of the posterior standard deviation in equation (2-5),

then
$$\sqrt{(n'/n'')\check{\mu}'} = s\sqrt{\check{\mu}'} \ .$$

Squaring both sides
$$\frac{n'}{n''}\check{\mu}' = s^2 \check{\mu}' \ .$$

Using equation (2-1), this may now be solved for n giving
$$n = \left(\frac{1}{s^2} - 1\right)n' \ , \qquad 0 < s \leqq 1 \ . \tag{2-7}$$

In essence, the above equation states that a sample of size n can be expected to reduce the prior standard deviation of $\tilde{\mu}$ by some factor s.

This has a certain amount of intuitive appeal. Notice that, if s=1, indicating that the prior standard deviation is satisfactory, the sample size is zero regardless of the value of n'. Additionally, if n' is interpreted to be the "weight" assigned the prior distribution, as suggested by Winkler [23], then as the prior distribution is given more weight in the analysis, the sample size increases. This is reasonable since the weight the analyst assigns to the prior distribution reflects his confidence in that distribution. Thus, if he has a great deal of confidence in the prior, it would take a large amount of sample information to significantly alter his beliefs.

The Solution Using a Bayesian Interval

Suppose that the decision maker would prefer to use some other measure of the posterior distribution. A reasonable measure would be a Bayesian prediction interval. The development of the solution procedure is similar to the previous section. A Bayesian prediction interval is an interval having a stated probability, e.g., $(1-\alpha)$, containing the variable of interest. In Figure 1, $\bar{\mu}''$ is the mean of the posterior distribution, a is the lower prediction limit, b is the upper prediction limit, and the shaded area represents the probability that $a \leqq \tilde{\mu} \leqq b$.

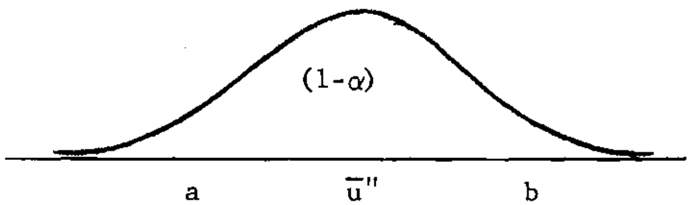

Figure 1. Generalized Bayesian Interval on $\tilde{\mu}''$

Now, assuming that the interval is centered on $\bar{\mu}''$, then the distance, d'', from a to b is given by [8]

$$d'' = 2t_{\alpha/2,\nu''}\sqrt{\check{\mu}''}\ . \tag{2-8}$$

Prior to sampling, the decision maker would be interested in the expected value of d'', which can be expressed as

$$E(d'') = 2t_{\alpha/2,\nu''}\, E(\sqrt{\check{\mu}''}\,).$$

Using equation (2-6) this becomes

$$E(d'') = 2t_{\alpha/2,\nu''}\sqrt{(n'/n'')\check{\mu}'}\,.$$

Squaring both sides gives

$$E(d'')^2 = 4\ t^2_{\alpha/2,\nu''}\ (n'/n'')\check{\mu}'\ .$$

This equation can then be solved for n yielding

$$n = \left[\frac{2t_{\alpha/2,\nu''}}{E(d'')}\right]^2 \check{\mu}'n' - n'\ . \tag{2-9}$$

The above equation parallels closely a result obtained by Cordova [6] where he shows that the minimum sample size required to establish a Bayesian interval of expected width k about the mean of a sampling distribution when the variance of that distribution, σ^2, is known is

$$n = \left(\frac{2Z_{\alpha/2}\sigma}{k}\right)^2 - n'\ .$$

He goes on to demonstrate that the quantity $(2Z_{\alpha/2}\sigma/k)^2$ is equivalent to the classical solution to the same problem.

Note that (2-9) cannot be solved explicitly for n. It is suggested that it be solved by the trial and error method. A good first approximation for n may be found by using $Z_{\alpha/2}$ in place of $t_{\alpha/2,\nu''}$ and solving for n.

Illustrating the Procedures

In this section, an example will be given of how each of the solution procedures may be applied in a realistic situation. It was decided to use the procedures in the context of an actual operational test. The test selected was OT II for the Lightweight Company Mortar System (LWCMS).

The LWCMS is being considered as a replacement for the M29 81mm mortar currently being used by the Army. The purpose of the test is to provide comparative data on the two types of mortars for assessing the relative operational performance and military utility of the LWCMS [14]. One of the MOE under consideration in this test is the time required for an individual to complete the gunner's examination. The gunner's examination is a practical test designed to measure how quickly an individual can perform certain essential operations in preparing a mortar to fire.

This MOE was previously examined during OT I. In that test, 14 individuals were given the gunner's exam using the 81mm mortar. They were then presented with two weeks of instruction on the LWCMS, after which they once more took the gunner's exam, this time using the LWCMS. The results of this test are contained in Appendix I. The format for the experiment in OT II is the same. The sample size problem is to determine the number of individuals to be used in that experiment. The first solution procedure to be illustrated will use the standard deviation as the measure on the posterior distribution.

The initial step in the procedure is to determine the value of the prior standard deviation of $\tilde{\mu}$. For notational purposes, the sample data relevant to the 81mm mortar will be denoted by X_{1i}, $i = 1,2, \ldots, 14$ and that associated with the LWCMS by X_{2i}, $i = 1,2, \ldots, 14$. To compute the value of $\sqrt{\breve{\mu}'}$ it is necessary to know n', v', and ν' of the prior distribution. Since this MOE was examined previously, the prior distribution for OT II may be equated to the posterior distribution of OT I. As pointed out in Chapter II, prior to OT I there is usually no internally generated data available; therefore, a diffuse prior distribution is appropriate. Thus, the posterior distributions associated with OT I are based solely on sample information. Considering this, the posterior parameters relative to OT I are given by [18, p. 298]

$$m'' = m \equiv \frac{\Sigma_i D_i}{n}$$

$$v'' = v \equiv \frac{\Sigma(D_i - m)^2}{n'' - 1}$$

$$n'' = n$$

$$\nu'' = n'' - 1 = n - 1$$

where

$$D_i = X_{1i} - X_{2i} \ .$$

Using the data from Appendix I, the following values were calculated:

$$n = 14$$

$$n'' = 14$$

$$m'' = 17.6 \text{ sec}$$

$$v'' = 2040.5 \text{ sec}^2$$

$$\nu'' = 13 \ .$$

The above values may now be used as the parameters of the prior distribution relative to OT II.

At this point it is possible to calculate an upper bound for the error induced by using equation (2-6) as an approximation for the expected standard deviation. With ν' equal to 13, this upper bound is six percent. If it is felt that a possible error of this magnitude is critical, then the approximation should not be used. It will be assumed here that such an error is not critical. The next step, then, is to calculate the value of the prior variance of $\tilde{\mu}$. Using equation (2-4)

$$\check{\mu}' = \left(\frac{2040.5}{14}\right)\left(\frac{13}{13-2}\right)$$

$$\check{\mu}' = 172.25 \text{ sec}^2 \ .$$

This produces a prior standard deviation of

$$\sqrt{\check{\mu}'} = 13.12 \text{ sec} \ .$$

The fact that this MOE is again being considered in OT II implies the above standard deviation is too large to formulate meaningful conclusions regarding $\tilde{\mu}$. What specific value of the posterior standard deviation would be acceptable is something which must be determined by the OTEA test designers. To assist in this decision, Table 2 depicts the sample sizes required to produce various expected values for the posterior standard deviation.

Table 2. Required Sample Sizes for Values of the Expected Posterior Standard Deviation (in seconds)

$E(\sqrt{\check{\mu}''})$	12.0	11.0	10.0	9.0	8.0	7.0	6.0	5.0	4.0	3.0	2.0	1.0
n	3	6	11	16	24	36	53	83	137	254	589	2396

The values of n were found by using equation (2-7) with

$$s = \frac{E(\sqrt{\check{\mu}''})}{13.12} \ .$$

All that remains is for the analyst to select the desirable value for the expected posterior standard deviation and obtain the required sample size from Table 2.

Now consider the solution procedure which uses a Bayesian interval on the posterior distribution as a measure. Based on the prior distribution, the length of an interval, centered on the mean, containing 90% of the probability is given by

$$\begin{aligned} d' &= 2t_{\alpha/2,\nu''}\sqrt{\check{\mu}'} \\ &= 2(1.761)(13.12) \\ &= 46.21 \text{ sec} . \end{aligned}$$

Suppose that it is desired to have the expected width of the Bayesian interval, with respect to the posterior distribution, be equal to

$$E(d'') = 20.00 \text{ sec} ,$$

then

$$E(d'')^2 = 400.00 \text{ sec}^2 .$$

Using equation (2-9)

$$n = \frac{(25_{.05,\nu''})^2(172.25)}{400}(14) - 14 .$$

To obtain a first approximation for n, $Z_{.05}$ is substituted for $t_{.05,\nu''}$ giving

$$n = \frac{4(1.645)^2(172.25)}{400}(14) - 14$$

$$n = 51.26 .$$

Rounding this up to the next greatest integer gives an initial value for n of 52. Using this sample size, n'' would equal 66, with the corresponding value of $t_{.05,65}$ being 1.6686. Substituting this value in equation (2-9) and solving for n gives

$$n = \frac{4(1.6686)^2(172.25)}{400}(14) - 14$$

$$n = 53.14 \; .$$

From this result it appears that the optimal n will lie somewhere between 52 and 54. Setting n equal to 53 and using the appropriate value for $t_{\alpha/2,\nu''}$ gives

$$n = \frac{4(1.6683)^2(172.25)}{400}(14) - 14$$

$$n = 53.12 \; .$$

Therefore, a sample of size 54 would reduce the expected width of a 90% Bayesian prediction interval to 20.

The procedures developed in this chapter have not considered any monetary constraints associated with the cost of sampling. In reality, however, such constraints play an important role in the determination of the optimal sample size. For example, it may be desirable to have an expected posterior standard deviation equal to one; however, the cost of having 2396 individuals involved in the experiment may be prohibitive. Therefore, it is necessary to achieve some balance between the E($\sqrt{\mu''}$) and the cost of sampling. This is the subject of the next chapter.

CHAPTER III

A DECISION THEORETIC APPROACH

Introduction

The problem which was examined in Chapter II will now be expanded to include economic considerations. The objective of this chapter will be to develop a procedure for determining the economically optimal sample size in the presence of monetary constraints. Initially, it appears that this problem could best be handled using Bayesian preposterior analysis. In Bayesian preposterior analysis, the decision maker examines each experiment available to him in an effort to determine if the expected value of the sample information is sufficient to justify the expected cost of the experiment. If more than one experiment satisfies this criterion, he then chooses the experiment which gives him the greatest expected net gain from sampling, where the expected net gain from sampling is equal to the expected value of sample information minus the expected cost of sampling [18].

The problem in this approach, however, is in describing the terminal utilities involved. The posterior distribution of $\tilde{\mu}$, with respect to any single MOE, is not used by itself as a basis for any terminal action. Rather, this distribution is considered along with the posterior distributions associated with many other MOE in the overall assessment procedure. Raiffa and Schlaifer have suggested a method for handling this type of problem [18].

> If this [sample information] is ultimately to be used in a number of second-stage action problems, then the immediate result of the experiment will presumably be a posterior distribution of $\tilde{\omega}$ [the variable of interest] in one guise or another. But how does this help to decide on sample size? One possible approach is to define some index of "tightness" of the posterior distribution (such as the variance if the distribution is symmetric) and then to decide on a substitution rate between the index of tightness and the cost of sampling.

It is a solution procedure similar to this which will be used here. Inherent in this approach is the idea of assigning an appropriate utility to each value of the index of tightness referred to above. Utility, as used here, is an expression of the relative worth to the decision maker of a particular value of that index.

The Solution Procedure

One logical choice for the index of tightness to be used in this study is the expected value of the posterior standard deviation. Assigning a utility to the values of $E(\sqrt{\tilde{\mu}''})$ may be difficult in the absolute sense. It may be considerably easier to assess the utilities relative to the total money available for testing, say K_t. For example, the decision maker may feel that an expected posterior standard deviation equal to one-half of the prior standard deviation may be worth one-half of K_t.

This last statement suggests the possibility of another index of tightness, that is the value of s, where s is the ratio of the expected posterior standard deviation to the prior standard deviation. This can be represented as a simpler function of n, which will be important later, and it seems to be more readily adaptable to assessing utilities. For these reasons, it will be used as the index of tightness for this study. Since s can assume any value on the interval (0,1], it will be necessary to define the utility of s using some continuous function, say U(s). Assume, for the moment, that this is possible.

The cost of sampling may also be thought of in terms of utility. If the cost of sampling is additive with fixed cost, K_f, and variable cost, K_r, then the total cost of sampling, K_s, may be represented by

$$K_s = K_f + K_r n \ . \tag{3-1}$$

The utility of the cost of sampling can be expressed as

$$U(K_s) = - K_s \ .$$

Then the utility of any experiment, e_n, where n refers to the sample size of the experiment, is given by

$$U(e_n) = U(s) - K_s \ . \tag{3-2}$$

Using the optimization criterion of maximizing utility, the solution to the sample size problem becomes one of finding the value of n which maximizes equation (3-2). What method may be used to find this optimum value of n clearly depends on the nature of U(s). Several different methods are examined in the following two sections.

Linear Utilities

In this section, the problem will be considered where U(s) is linear with respect to s. In this case

$$U(e_n) = as+b - K_s \ . \tag{3-3}$$

Now, equation (2-7) may be solved for s in terms of n and n' yielding

$$s = (n')^{1/2}(n'+n)^{-1/2}. \tag{3-4}$$

Using the above expression for s and equation (3-1) for K_s, equation (3-3) may be written

$$U(e_n) = a(n')^{1/2}(n'+n)^{-1/2}+b - K_f - K_r n \ .$$

Differentiating with respect to n gives

$$\frac{dU(e_n)}{dn} = a(n')^{1/2}(-1/2)(n'+n)^{-3/2} - K_r$$

$$= - (a/2)(n')^{1/2}(n'+n)^{-3/2} - K_r \ .$$

It does not appear that this can be readily optimized by equating the first derivative to zero and solving for n. Perhaps an easier way would be to use a nonlinear one-dimensional search such as the golden section search. To do this, the objective function must be unimodal and continuous [3]. One way of checking for these properties is to investigate the convexity, or concavity, of the function. If the function is either convex or concave, the properties of unimodality and continuity are guaranteed [3]. This will be done by examining the sign of the second derivative.

$$\frac{d^2U(e_n)}{dn^2} = -\ (a/2)(n')^{1/2}(-3/2)(n'+n)^{-5/2}$$

$$= (3a/4)(n')^{1/2}(n'+n)^{-5/2}\ .$$

Since n' and n are both greater than or equal to zero, the sign of the second derivative depends solely on the sign of a. The value of a represents the slope of U(s). If the slope is positive, the decision maker is expressing a greater preference for larger values of s than for smaller ones. This does not seem to be reasonable. Therefore, it is felt that a can be required to be less than zero with no loss of generality. With this restriction, the function is concave and the golden section method is applicable. A computer program designed to solve this problem is contained in Appendix II.

Power Function Utilities

Now suppose that U(s) is of the form

$$U(s) = (1-s)^c\ K_t\ , \tag{3-5}$$

then

$$U(e_n) = (1-s)^c\ K_t - K_s\ .$$

Substituting for s and K_s gives

$$U(e_n) = [1 - (n')^{1/2}(n'+n)^{-1/2}]^c\ K_t - K_f - K_r n\ .$$

Differentiating once with respect to n gives

$$\frac{dU(e_n)}{dn} = cK_t[1-(n')^{1/2}(n'+n)^{-1/2}]^{c-1}(-1)(n')^{1/2}(-1/2)(n'+n)^{-3/2} - K_r$$

$$= \frac{cK_t}{2}(n')^{1/2}[1-(n')^{1/2}(n'+n)^{-1/2}]^{c-1}(n'+n)^{-3/2} - K_r \ .$$

Once more it appears that the golden section method may be the best solution approach. Checking for concavity again involves taking the second derivative.

$$\frac{d^2U(e_n)}{dn^2} = (n')^{1/2}\left(\frac{cK_t}{2}\right)\left\{\left[1-(n')^{1/2}(n'+n)^{-1/2}\right]^{c-1}(-3/2)(n'+n)^{-5/2}\right. \quad (3\text{-}6)$$

$$+ (n'+n)^{-3/2}(c-1)[1-(n')^{1/2}(n'+n)^{-1/2}]^{c-2}(-1)(n')^{1/2}$$

$$\left.\times (-1/2)(n'+n)^{-3/2}\right\}$$

$$= (n')^{1/2}\left(\frac{cK_t}{2}\right)\left\{(-3/2)[1-(n')^{1/2}(n'+n)^{-1/2}]^{c-1}(n'+n)^{-5/2}\right.$$

$$\left.+ (c-1)(1/2)(n')^{1/2}[1-(n')^{1/2}(n'+n)^{-1/2}]^{c-2}(n'+n)^{-3}\right\} \ .$$

Determining the sign of the second derivative, using equation (3-6), is not as straightforward as in the case of the linear utility functions. Clearly, this sign depends on the sign of the parameter c, but also it depends on the relative magnitudes of the two terms inside the brackets. First, consider values of c such that $0 < c \leqq 1$. This makes both terms inside the brackets negative while the term outside the brackets is positive, thereby making the sign of the entire expression negative.

Thus, for these values of c, the function is concave. Figure 2 shows some typical members of this family of functions.

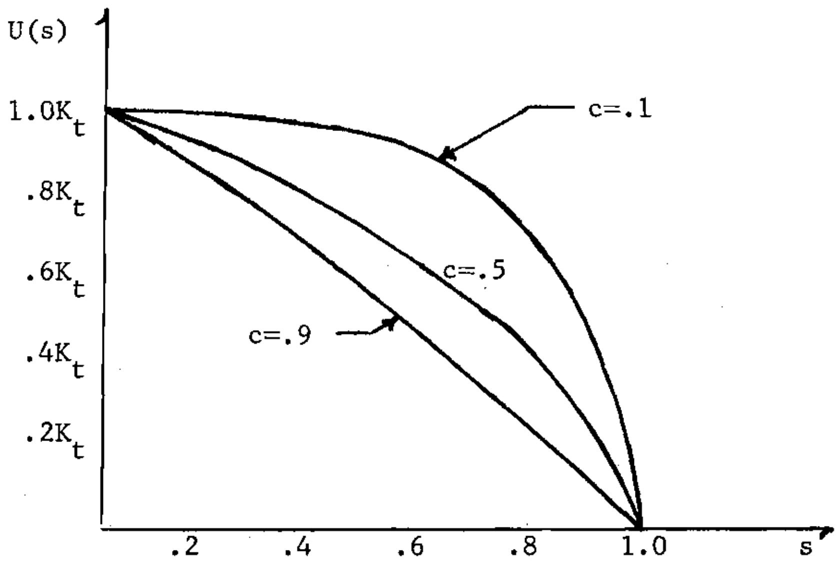

Figure 2. Typical Power Function Utilities for $0 < c \leqq 1$

Next, consider values of c which are less than zero. In this case, the factor outside the brackets, as well as the terms inside the brackets, are negative. Therefore, the function will be convex for all negative values of c. Note that, if c is negative, equation (3-5) may be rewritten as

$$U(s) = \left(\frac{1}{1-s}\right)^q K_t$$

where q is equal to minus c and is, therefore, a positive quantity. It is obvious from this equation that U(s) increases as the value of s increases. This would indicate that the decision maker has a greater preference for large values of s than he does for small ones. Such a preference would be inconsistent with the objectives of the testing and, therefore, will not be considered in this study.

Finally, for values of c greater than one, the family of utility functions would be comprised of functions similar to those shown in Figure 3.

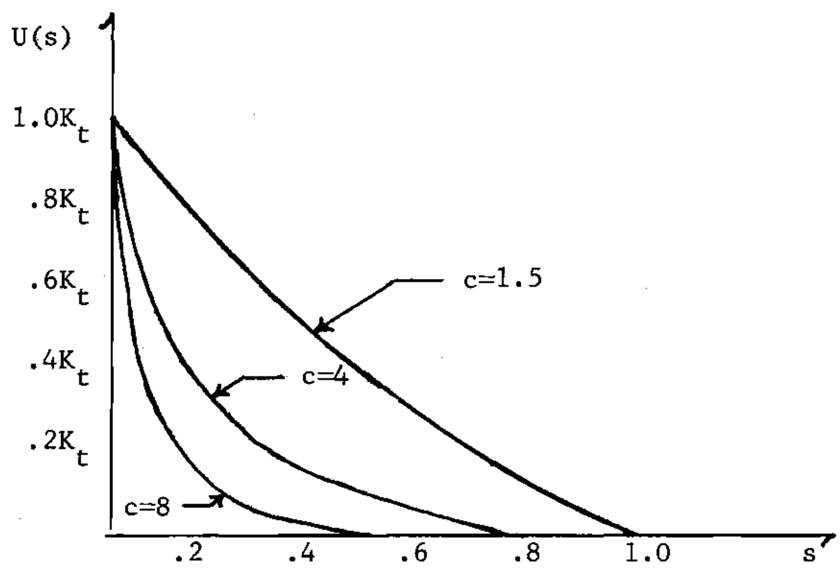

Figure 3. Typical Power Function Utilities for $c > 1$

Functions such as these could be very useful since they would allow the decision maker to express a greater preference for small values of s in comparison to larger values. Unfortunately, no general statement can be made as to whether they are unimodal or not. This does not mean, however, that they cannot be optimized. Bazaraa and Shetty [3] suggest that the golden section search may still be used for non-unimodal functions by subdividing the interval of uncertainty into a number of smaller intervals and searching over these smaller intervals. The object is to select the size of the smaller intervals so that it is reasonably certain that the function is unimodal over that interval. This procedure will be illustrated in the next section.

Illustrating the Procedure

The solution procedure will now be illustrated using both types of utility functions described previously. Before doing this, however, the computer program used in the analysis will be discussed briefly. This program is shown in Appendix II. It is designed to perform a golden section search using either the linear or power function utilities. The golden section method determines the optimum value for unimodal, continuous functions by successively reducing the size of an interval of uncertainty [3]. Therefore, it is necessary to initially determine some interval of values for the sample size which is believed to contain the optimal value. This is done in the program by setting the lower bound of the interval equal to zero and the upper bound equal to the largest sample size allowable under the existing budget constraint. If the function is unimodal, this interval is then systematically reduced until it is less than or equal to one. If the function is not unimodal, the decision maker is then required to subdivide the interval of uncertainty and then each of the smaller intervals is searched.

The same experiment used in Chapter II will be used for this illustration. In order to do this, however, several additional inputs are necessary, specifically, the budget constraint, K_t, the sampling costs, K_f and K_r, and the utility function, U(s).

To think of a budget constraint and a cost of sampling associated with a single MOE may be somewhat unrealistic. In practice, a single experiment will produce data on many different MOE. Most of the time, the only budget and cost figures associated with the test are aggregate amounts in the form depicted in Table 3. Therefore, rather than attempting to determine the sampling cost for a specific MOE and the total money available for testing that MOE, it may be much more realistic to allocate to each MOE some proportion of the aggregate budget and estimated costs. This is not currently being done, so it will be necessary to approximate these values.

Table 3. Total Cost Estimates (Direct Costs) [14]

Elements of Cost	Estimated Cost (In Thousands of Dollars)
1. Test Directorate Operating Costs	19.1
2. Player Participants	22.1
3. Test Facilities	30.0
4. Items to Be Tested	.5
5. Data Collection, Processing and Analysis	6.4
6. Ammunition	145.4
7. Pre-Test Training	2.1
8. Photographic Support	15.0
9. Other Costs	4.5
Total	245.1

It is suggested that the proportion of the aggregate budget to be assigned to a specific MOE be commensurate with that MOE's relative importance. The OTEA already assesses the relative importance of MOE in qualitative terms [11]. All that is required then is to quantify this assessment, perhaps through a series of weighting functions. It is not anticipated that this requirement would represent a major problem to OTEA test design personnel who have detailed information on the relationship between the data requirements and the operational issues being examined.

Since this type of information is not presently available, a very simplistic approach was taken to the allocation problem. Each of the MOE was weighted equally in determining the individual budget constraint. Based on an imposed test budget constraint of $250,000.00, the individual budget constraint for each MOE, K_t, was derived to be $1,724.00.

The derivation of values for the fixed and variable costs was accomplished in a slightly different manner. The aggregate estimated fixed cost was defined to be the sum of all those costs in Table 3 except the costs of player participants and ammunition. This resulted in a total figure of $77,600.00. This figure was then divided by the length of the test in weeks to yield a fixed cost per week of $5,969.00. Using this weekly cost estimate, each phase of the test was assigned a fraction of the total estimated fixed cost based on the time required to conduct that particular phase. The fixed cost associated with each phase was then distributed equally among the MOE being examined in that phase. Table 4 presents the results of this process.

The variable costs are of two types, those associated with a sample size requirement for a certain number of <u>different</u> individuals and those associated with the requirement for the expenditure of a specified number of rounds of ammunition. Both of these variable costs were approximated by dividing the appropriate total estimated cost figures presented in Table 4 by the total estimated requirements for that resource [14]. This resulted in a variable cost for personnel of \$57.00 per week per man and a cost of ammunition of \$13.00 per round.

Table 4. Allocation of Estimated Fixed Costs

Phase	Length of Phase (weeks)	Fixed Cost for Phase (\$)	No. MOE Examined	Fixed Cost per MOE (\$)
1. Training	2	11,938	28	426
2. Pilot Test	1	5,969	0	0
3. Field Exercise	3	17,908	73	245
4. Live Fire	6	35,815	36	995
5. Parachute Delivery Demonstration	1	5,969	8	746

The MOE of interest in this illustration is to be examined during the training phase so the fixed cost, K_f, is \$426.00. The test design calls for using the same number of individuals throughout the training phase. Therefore, the variable cost, K_r, was derived by multiplying the cost per man per week by the number of weeks required to complete the training phase and then dividing the result by the number of MOE examined during this phase. This process resulted in a value of \$4.00 for K_r.

The above methods for approximating budget constraints and sampling costs are not necessarily being advocated for use by OTEA; they were used here to provide a starting point for the demonstration. This being accomplished, it remains to select an appropriate function for U(s).

The first case to be considered is that of a linear utility function. The form of this function is

$$U(s) = as + b \qquad \begin{array}{c} a \leqq 0 \\ 0 < s \leqq 1 \end{array}$$

Consider Figure 4 below, by varying the values of the parameters a and b, it is possible to represent U(s) by any negatively sloped straight line which intersects the s-axis between zero and one. This provides the decision maker with a rich family of linear functions from which to choose. The one chosen for this illustration is the one depicted in Figure 4.

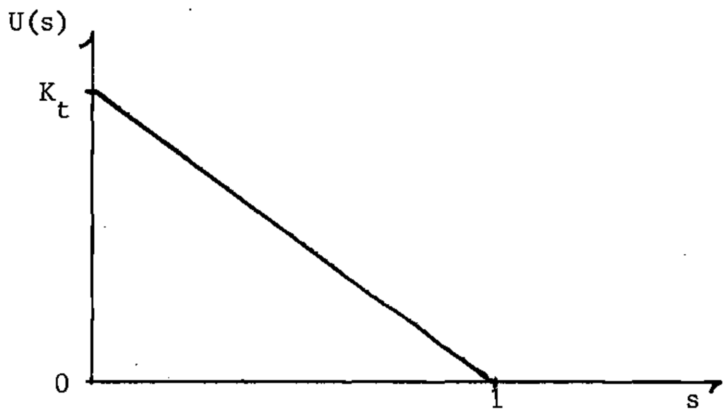

Figure 4. Linear Utility Function

The equation for this function is

$$U(s) = -K_t s + K_t = K_t(1-s)$$

Using this utility function and the budget constraint and sampling cost previously derived, the objective function becomes

$$U(e_n) = K_t[1-(n')^{-1/2}(n'+n)^{-1/2}] - K_f - K_r n \ .$$

This objective function was used in the computer analysis, the results of which are presented in Table 5. The economically optimal sample size is 73. This will reduce the magnitude of the prior standard deviation by a factor of approximately four-tenths.

Table 5. Computer Analysis Using Linear Utility

Lower Limit	Upper Limit	N1	N2	U(N1)	U(N2)
0.00	324.50	123.93	200.54	.253	.055
0.00	200.54	76.61	123.93	.314	.253
0.00	123.93	47.33	76.61	.285	.314
47.33	123.93	76.61	94.66	.314	.301
47.33	94.66	65.38	76.61	.312	.314
65.38	94.66	76.61	83.43	.314	.311
65.38	83.43	72.19	76.61	.314	.314
65.38	76.61	69.79	72.19	.314	.314
69.79	76.61	72.19	74.20	.314	.314
69.79	74.20	71.79	72.19	.314	.314
71.79	74.20	72.19	73.80	.314	.314
71.79	73.80	72.19	73.40	.314	.314
72.19	73.80	72.60	73.40	.314	.314

This same analysis will now be conducted using two power function utilities. The first will be defined by

$$U(s) = (1-s)^{1/2} K_t \qquad 0 < s \leqq 1$$

Using this utility, the objective function is

$$U(e_n) = (1-s)^{1/2} K_t - K_f - K_r n \qquad 0 < s \leqq 1$$

This function was entered into the computer program giving the results shown in Table 6. As seen from this table, the economically optimal sample size is 52. This is a smaller sample size than obtained by using the linear utility function. This result is to be expected since this power function gives more weight to larger values of s.

Table 6. Computer Analysis Using Power Function with c = 1/2

Lower Limit	Upper Limit	N1	N2	U(N1)	U(N2)
0.00	324.50	123.93	200.54	.501	.259
0.00	200.54	76.61	123.93	.611	.501
0.00	123.93	47.33	76.61	.631	.611
0.00	76.61	29.28	47.33	.589	.631
29.28	76.61	47.33	58.56	.631	.631
47.33	76.61	58.56	65.38	.631	.625
47.33	65.38	54.15	58.56	.632	.631
47.33	58.56	51.74	54.15	.632	.632
47.33	54.15	49.73	51.74	.632	.632
49.73	54.15	51.74	52.14	.632	.632
51.74	54.15	52.14	53.74	.632	.632
51.74	53.74	52.14	53.34	.632	.632
51.74	53.34	52.14	52.94	.632	.632
51.74	52.94	52.14	52.54	.632	.632

The second power function utility to be considered has the parameter c equal to 1.5. This function is shown in Figure 3. Since this particular function is not guaranteed to be unimodal over all n, the method of subdividing the interval of uncertainty into a number of smaller intervals was employed. The interval of uncertainty, based on the budget constraint, is (0.00, 324.50). This interval was searched using subintervals of length 20. The results are shown in Table 7. As can be seen from this table, the optimal sample size is 83. Note that the utility of the experiment steadily increases until the optimal sample size is reached and then steadily declines over the remaining values of n. Thus, it is reasonably certain that a sample of size 83 is, in fact, a global optimal.

Table 7. Results of Computer Analysis Using Power Function Utility with c = 1.5

Subinterval	Optimal Sample Size for Subinterval	Utility of Experiment
0 - 20	20	-.144
20 - 40	40	.004
40 - 60	60	.065
60 - 80	80	.083
80 - 100	83	.084
100 - 120	100	.076
120 - 140	120	.053
140 - 160	140	.019
160 - 180	160	.022
180 - 200	180	-.069
200 - 220	200	-.121
220 - 240	220	-.176
240 - 260	240	-.234
260 - 280	260	-.356
300 - 320	300	-.420

CHAPTER IV

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The procedures developed in this study have been structured around Bayesian inference and decision theory. The demonstration of these procedures has been confined to instances where the prior distributions have been developed using only objective prior data. This was done to show that the procedures could be beneficial regardless of the strict limitations imposed on the use of any subjective prior information. Considered in this context, however, the analysis is not Bayesian in the purest sense, nor would such an approach be appropriate when the use of subjective information is disallowed.

Nonetheless, it is felt that the inference approach presented in Chapter II represents a viable solution to the sample size determination problem currently faced by OTEA test designers. Its major advantage over the presently employed procedures is that it has been specifically designed to address the problem of making statistically valid inferences regarding the <u>difference</u> in two random variables.

The decision theoretic approach, described in Chapter III, provides the decision maker with a quantitative procedure for comparing the expected results of an experiment to the cost of that experiment. If the decision maker is willing to accept the concepts of utility theory, and to apply them, this procedure will provide the economically optimal sample size.

Finally, should the limitations imposed on the use of subjective information be relaxed, the methodology presented in this study could be used in conducting a completely Bayesian analysis.

Recommendations

The greatest limitation to the methodology developed in this study is that it is applicable only to the case of sizing an experiment for a single MOE. The logical extension of this is to the case of multiple MOE. There are at least two approaches to analyzing this case. One would be to apply multivariate Bayesian statistical theory combined with multidimensional nonlinear programming algorithms. A second approach would be to view the money required to perform each of the experiments involved in an operational test as a capital investment and the utility of each of the experiments as the return on that investment. Formulated in this manner the problem might be solved utilizing capital budgeting techniques. If it is possible to extend the methodology to include multiple MOE, then it may be possible to use it in multifactor experimental design problems.

Aside from extending the methodology, several other areas warrant further investigation. First, is the assumption that the normal process may be used as a reasonable model for a large number of operational testing problems. Closely associated with this would be an investigation of the variation in results when the sampling process is not normal.

As a final recommendation, it is suggested that the procedures outlined in this study be utilized in designing a number of operational tests and that these results be compared to the results obtained using the presently employed methods.

APPENDIX I

LIGHTWEIGHT COMPANY MORTAR SYSTEM OT I TEST DATA

Gunner's Examination Times [13]

Test Participant	System: 81 mm (sec)	System: LWCMS (sec)	Difference in Performance
1	358.0	303.4	54.6
2	367.0	350.8	16.2
3	299.0	330.0	- 31.0
4	261.0	147.5	113.5
5	380.0	313.0	67.0
6	226.8	250.0	- 23.2
7	272.0	247.0	25.0
8	239.8	273.0	- 33.2
9	235.0	258.0	- 23.0
10	247.5	244.8	2.7
11	279.1	242.7	36.4
12	303.0	234.2	68.8
13	240.9	250.7	- 9.8
14	279.0	296.9	- 17.9

APPENDIX II

FORTRAN PROGRAM FOR THE GOLDEN SECTION SEARCH TECHNIQUE

```
      PROGRAM MAIN(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
C
C     THIS PROGRAM PERFORMS A NONLINEAR,
C     ONE-DIMENSIONAL SEARCH USING THE GOLDEN
C     SECTION METHOD.
C
C
C     LIST OF VARIABLES.
C
C     AKT        - TOTAL MONEY ALLOCATED TO TESTING
C                  THIS MOE (BUDGET CONSTRAINT).
C     AKF        - FIXED COST OF TESTING.
C     AKR        - VARIABLE COST OF TESTING.
C     NP         - WEIGHT GIVEN TO PRIOR INFORMATION.
C     XNP        - TRANSFORM ON NP TO MAKE IT A
C                  REAL VARIABLE.
C     IP         - INDICATOR VARIABLE USED TO DESIGNATE
C                  TYPE OF UTILITY FUNCTION.
C     A,B        - SLOPE AND INTERCEPT OF LINEAR UTILITY
C                  FUNCTION.
C     C          - EXPONENT OF POWER UTILITY FUNCTION.
C     K          - COUNTER DESIGNATING NUMBER OF
C                  ITERATIONS USED IN THE SEARCH.
C     SUBLIM     - LOWER LIMIT ON THE INTERVAL OF
C                  UNCERTAINTY.
C     UPLIM      - UPPER LIMIT ON THE INTERVAL OF
C                  UNCERTAINTY.
C     XN1,XN2    - POINTS ON THE INTERVAL OF UNCERTAINTY
C                  BEING EXAMINED.
C     UE1        - UTILITY OF EXPERIMENT OF SIZE
C                  XN1.
C     UE2        - UTILITY OF EXPERIMENT OF SIZE
C                  XN2.
C     ICHECK     - VARIABLE USED TO CHECK FOR
C                  TERMINATION.
C
C
      DIMENSION SUBLIM(20),UPLIM(20),UE1(20),UE2(20),
     *XN1(20),XN2(20),VP(2)
      REAL MP(2)
      INTEGER IANS,KANS,JANS
  100 FORMAT(///*ENTER THE TOTAL AMOUNT OF MONEY WHICH  */
     **YOU WISH TO ALLOCATE TO TESTING THIS MOE.*/)
  101 FORMAT(///*ENTER THE SET-UP AND REPLICATION COSTS IN*/
     ** THAT ORDER.*/)
  102 FORMAT(///*ENTER WEIGHT OF PRIOR INFORMATION.*/)
  103 FORMAT(///*ENTER TYPE OF UTILITY FUNCTION TO BE USED*/
     **(ENTER 1 FOR LINEAR AND 2 FOR POWER).*/)
  104 FORMAT(///*ENTER SLOPE AND INTERCEPT, *,
     **IN THAT ORDER*/)
```

```
  105 FORMAT(///*ENTER EXPONENT.*/)
  107 FORMAT(A6)
  200 FORMAT(I2,5X,F7.2,5X,F7.2,5X,F7.2,5X,F7.2,5X,2F8.3)
  201 FORMAT(///9X,*LOWER*,6X,*UPPER*/9X,*LIMIT*,
     *6X,*LIMIT*,9X,*N1*,10X,*N2*,10X,*U(N1)*,
     *3X,*U(N2)*)
  202 FORMAT(///*IS THE OBJECTIVE FUNCTION UNIMODAL?*/)
  203 FORMAT(///*THE INTERVAL OF UNCERTAINTY IS (*,
     *F7.2,*,*,F7.2,*)*/)
  204 FORMAT(///*ENTER LOWER AND UPPER BOUNDS*,
     ** FOR SUBINTERVAL.*/)
  205 FORMAT(///*DO YOU WISH TO SEARCH ANOTHER *,
     **SUBINTERVAL?*/)
      DATA KANS/6HYES    /
C
C     ENTER THE BUDGET CONSTRAINT.
C
      WRITE(6,100)
      READ(5,*) AKT
C
C     ENTER ESTIMATED COSTS.
C
      WRITE(6,101)
      READ(5,*) AKF,AKR
C
C     ENTER WEIGHT OF PRIOR INFORMATION.
C
      WRITE(6,102)
      READ(5,*) NP
      XNP=NP
C
C     DETERMINE TYPE OF UTILITY FUNCTION TO BE USED.
C
      WRITE(6,103)
      READ(5,*) IP
C
C     READ UTILITY FUNCTION PARAMETERS.
C
      IF(IP.EQ.2) GO TO 10
      WRITE(6,104)
      READ(5,*) A,B
      GO TO 20
   10 WRITE(6,105)
      READ(5,*) C
C
C     SET UP INITIAL INTERVAL OF UNCERTAINTY.
C
   20 K=1
      SUBLIM(K)=0
      UPLIM(K)=(AKT-AKF)/AKR
```

```
C
C     SEE IF THE FUNCTION IS UNIMODAL.
C
      WRITE(6,202)
      READ(5,107) IANS
      IF(IANS.EQ.KANS) GO TO 501
C
C     DETERMINE WHAT SUBINTERVAL IS TO BE SEARCHED.
C
      WRITE(6,203) SUBLIM(K),UPLIM(K)
  502 K=1
      WRITE(6,204)
      READ(5,*) SUBLIM(K),UPLIM(K)
C
C     CALCULATE INITIAL VALUES FOR XN1 AND XN2.
C
  501 XN1(K)=.381924*(UPLIM(K)-SUBLIM(K))+SUBLIM(K)
      XN2(K)=.618*(UPLIM(K)-SUBLIM(K))+SUBLIM(K)
C
C     WRITE COLUMN HEADINGS.
C
      WRITE(6,201)
C
C     CALCULATE UTILITIES.
C
  920 R1=XNP/(XNP+XN1(K))
      S1=SQRT(R1)

      R2=XNP/(XNP+XN2(K))
      S2=SQRT(R2)
      IF(IP.EQ.1) GO TO 40
      UE1(K)=(1.0-S1)**C*AKT-AKF-AKR*XN1(K)
      UE2(K)=(1.0-S2)**C*AKT-AKF-AKR*XN2(K)
      GO TO 21
   40 UE1(K)=A*S1+B-AKF-AKR*XN1(K)
      UE2(K)=A*S2+B-AKF-AKR*XN2(K)
   21 IF(XN1(K).LE.XN2(K)) GO TO 30
      A1=XN1(K)
      A2=XN2(K)
      XN1(K)=A2
      XN2(K)=A1
C
C     WRITE RESULTS.
C
   30 WRITE(6,200) K,SUBLIM(K),UPLIM(K),XN1(K),XN2(K),
     *UE1(K),UE2(K)
C
C     COMPARE UTILITIES.
C
      IF(UE1(K).LE.UE2(K)) GO TO 934
```

```
C
C
C     SET UP NEW INTERVAL OF UNCERTAINTY.
C
      UPLIM(K+1)=XN2(K)
      SUBLIM(K+1)=SUBLIM(K)
      XN2(K+1)=XN1(K)
      XN1(K+1)=UPLIM(K+1)-XN2(K+1)+SUBLIM(K+1)
      GO TO 944
  934 UPLIM(K+1)=UPLIM(K)
      SUBLIM(K+1)=XN1(K)
      XN1(K+1)=XN2(K)
      XN2(K+1)=UPLIM(K+1)-XN1(K+1)+SUBLIM(K+1)
C
C     CHECK FOR TERMINATION.
C
  944 ICHECK=SUBLIM(K)+2
      IF(ICHECK.GE.UPLIM(K).OR.K.GT.20) GO TO 930
      K=K+1
      GO TO 920
  930 CONTINUE
      IF(IANS.EQ.KANS) GO TO 931
C
C     SEE IF ANOTHER SUBINTERVAL IS TO BE SEARCHED.
C
      WRITE(6,205)
      READ(5,107) JANS
      IF(JANS.EQ.KANS) GO TO 502
  931 CONTINUE
      STOP
      END
```

APPENDIX III

EXPLANATION OF NOTATION

Chapter I

μ	mean of normal density function
σ^2	variance of normal density function
$\overline{x}$	sample mean
s^2	sample variance
$f'(\theta)$	prior distribution of $\tilde{\theta}$
$f(y\|\theta)$	likelihood function for $\tilde{y}$ given θ
$f''(\theta\|y)$	posterior distribution of $\tilde{\theta}$

Chapter II

$f_{N\gamma}(\tilde{\mu},\tilde{h}\|m',v',n',\nu')$	normal-gamma density function
h	inverse of σ^2
m',v',n',ν'	prior parameters for a normal-gamma density function (these are interpreted on page 11)
m'',v'',n'',ν''	posterior parameters for a normal-gamma density function (these are defined mathematically on page 12)
m,v,n,ν	parameters of a normal sampling distribution (these are defined mathematically on page
$f_S(\mu\|m,n/v,\nu)$	density function for Student's t distribution
$\overline{\mu}$	expected value of $\tilde{\mu}$
$\check{\mu}$	variance of $\tilde{\mu}$
$\check{\mu}'$	prior variance of $\tilde{\mu}$

$\sqrt{\check{\mu}'}$ prior standard deviation of $\tilde{\mu}$

$\bar{\mu}'$ prior mean value of $\tilde{\mu}$

$\sqrt{\check{\mu}''}$ posterior standard deviation

s ratio of the expected posterior standard deviation to the prior standard deviation

$\check{\mu}''$ posterior variance of $\tilde{\mu}$

$\bar{\mu}''$ posterior mean value of $\tilde{\mu}$

d'' length of a $(1-\alpha)$ Bayesian prediction interval on the posterior distribution

d' length of a $(1-\alpha)$ Bayesian prediction interval on the prior distribution

Chapter III

K_t total money allocated to testing

K_s total cost of sampling

K_f fixed cost of sampling

K_r variable cost of sampling

U(s) utility function for s

$U(K_s)$ utility function for the cost of sampling

e_n experiment with sample size of n

$U(e_n)$ utility function for an experiment of sample size n

BIBLIOGRAPHY

1. Anscombe, F. J., "A Definition of Subjective Probability," Ann. Math. Stat., Vol. 34, 1963, pp. 199-205.

2. AR 10-4, dated 30 December 1974, Organization and Functions of U.S. Army Operational Test and Evaluation Agency, Headquarters, Department of the Army, Washington, D. C.

3. Bazaraa, M. S., and Shetty, C. M., Nonlinear Programming, Unpublished Manuscript, Georgia Institute of Technology, Atlanta, Georgia, 1975.

4. Birnbaum, A., "On the Foundations of Statistical Inference," Journal of the American Statistical Association, Vol. 57, 1962, pp. 269-306.

5. Chernoff, H., and Moses, L. E., Elementary Decision Theory, John Wiley and Sons, Inc., New York, 1959.

6. Cordova, R. L., An Application of Bayesian Analysis in Determining Appropriate Sample Sizes for Use in U.S. Army Operational Tests, Unpublished Master's Thesis, Georgia Institute of Technology, Atlanta, Georgia, 1975.

7. De Groot, M. H., Optimal Statistical Decisions, McGraw-Hill Book Co., New York, 1970.

8. Hines, W. W., and Montgomery, D. C., Probability and Statistics in Engineering and Management Science, The Ronald Press Co., New York, 1972.

9. Lilliefors, H., "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," Journal of the American Statistical Association, Vol. 62, 1967, pp. 399-402.

10. Miller, J. R., Professional Decision Making, Praeger Publishers, New York, 1970.

11. Operations Research Associates, "Techniques and Guidelines," Operational Test Methodology Guide, Vol. II, January 1974.

12. OTEA, "Test Design Plan for Lightweight Company Mortar System (LWCMS)," U.S. Army Operational Test and Evaluation Agency, Ft. Belvoir, Virginia, 3 August 1973.

BIBLIOGRAPHY (Concluded)

13. OTEA, "Lightweight Company Mortar System (LWCMS) Operational Test I (OT I) Final Report," U.S. Army Operational Test and Evaluation Agency, Ft. Belvoir, Virginia, September 1973.

14. OTEA, "Lightweight Company Mortar System, Operational Test II, Test Design Plan," U.S. Army Operational Test and Evaluation Agency, Falls Church, Virginia, December 1975.

15. Plackett, R. L., "Current Trends in Statistical Inference," J. R. Statist. Soc. A, Vol. 129, 1966, pp. 249-267.

16. Pratt, J. W., Raiffa, H., and Schlaifer, R., Introduction to Statistical Decision Theory, McGraw-Hill Book Co., New York, 1965.

17. Raiffa, H., Decision Analysis, Addison-Wesley Publishing Co., Reading, Mass., 1970.

18. Raiffa, H., and Schlaifer, R., Applied Statistical Decision Theory, The MIT Press, Boston, 1961.

19. Savage, L. J., "Reading Suggestions for the Foundations of Statistics," American Statistician, Vol. 24, 1970, pp. 23-27.

20. Savage, L. J., "Elicitation of Personal Probabilities and Expectations," Journal of the American Statistical Association, Vol. 66, 1971, pp. 783-801.

21. Winkler, R. L., "The Assessment of Prior Distributions in Bayesian Analysis," Journal of the American Statistical Association, Vol. 62, 1967, pp. 776-800.

22. Winkler, R. L., "Probabilistic Predictions: Some Experimental Results," Journal of the American Statistical Association, Vol. 66, 1971, pp. 675-685.

23. Winkler, R. L., Introduction to Bayesian Inference and Decision, Holt, Rinehart and Winston, Inc., New York, 1972.